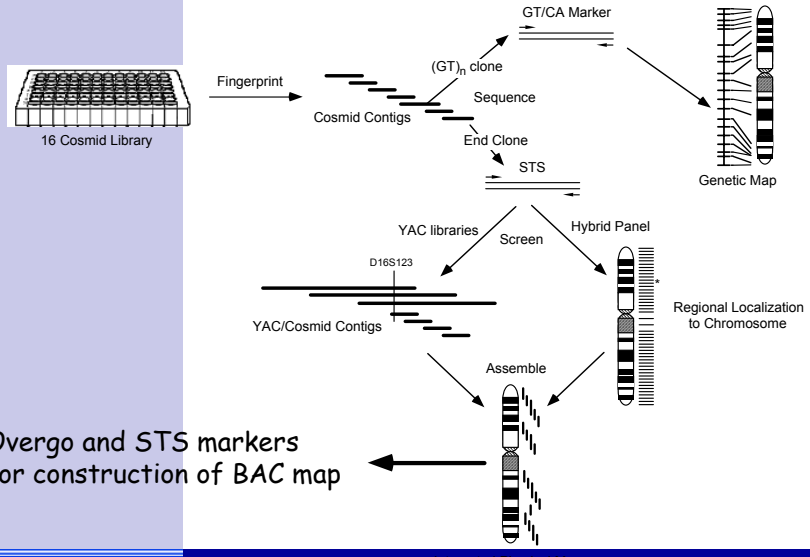
DOE JOINT GENOME INSTITUTE

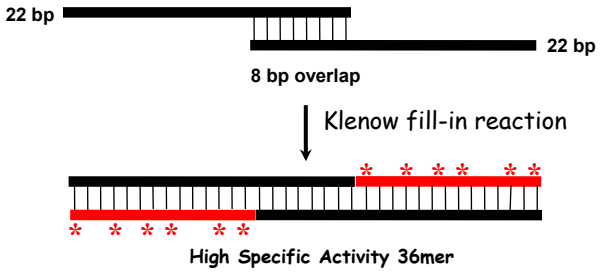
YAC/STS/Cosmid Map of Chromosome 16



Overgo and STS markers
for construction of BAC map

DOE JOINT GENOME INSTITUTE

Overgo Probe

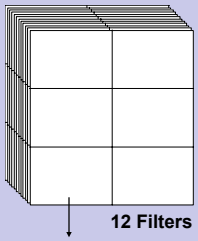


Features

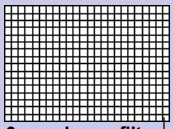
1. Easily designed to be unique
2. Simple to handle
3. Suitable for complex pooled hybridizations
4. High specific activity labeling

DOE JOINT GENOME INSTITUTE

Top pool hybridization



12 Filters

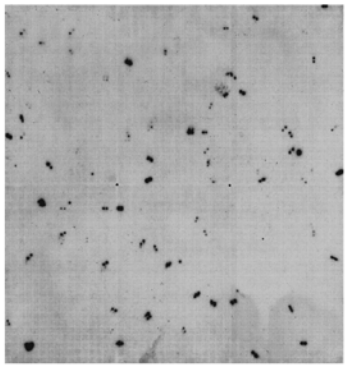


6 panels per filter

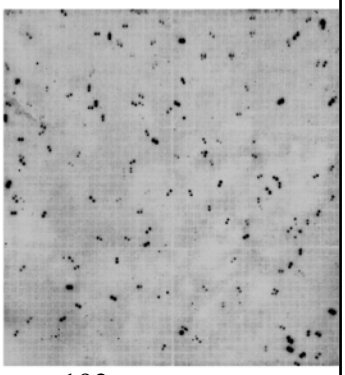
24 x 16 blocks of 4x4 colonies

221,184 BAC clones

Pod name	1	2-9	3-4	5-6	7-8-10	Total
Number of STSs	96	192	192	192	236	908



96 overgos



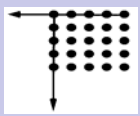
192 overgos

DOE JOINT GENOME INSTITUTE

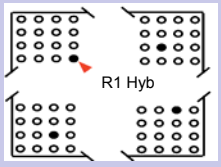
Two dimensional hybridization

R1

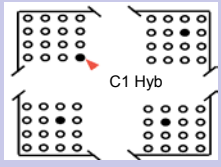
Overgo Row & Column pools



C1

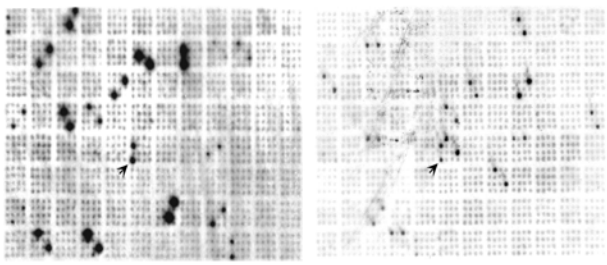


R1 Hyb



C1 Hyb

Group	YAC	BRPT	WI	SHGC	C16	Total
Number of STSs	402	128	166	64	148	908
Column pools	24	24	24	24	24	120
Row pools	17	6	8	3	7	41

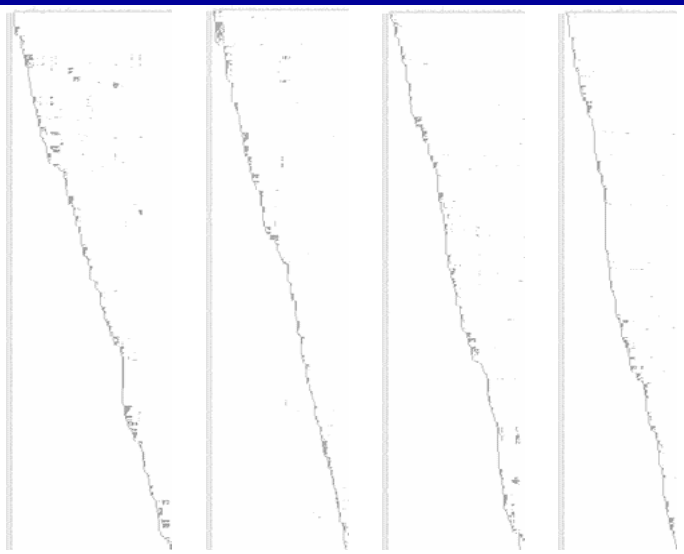


Left: Pool Column 17 of WIC group; filter number 5. Right: Pool Row 6 of WIC group; filter number 5. The overgo at the intersection of the two pools is WI-14879.

US DOE Joint Genome Institute

DOE JOINT GENOME INSTITUTE

Probe-content Maps for 16q

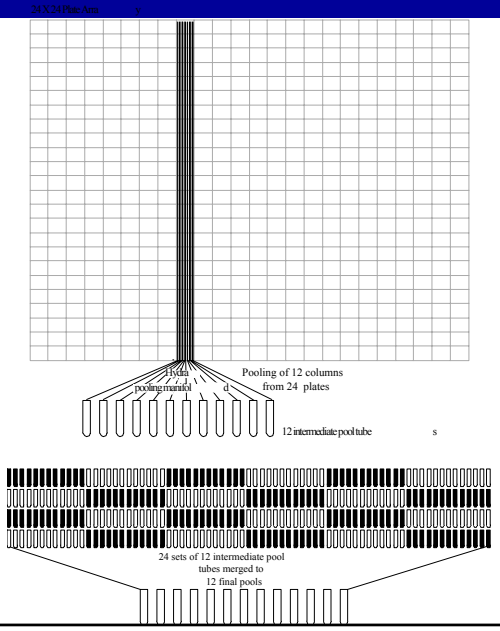


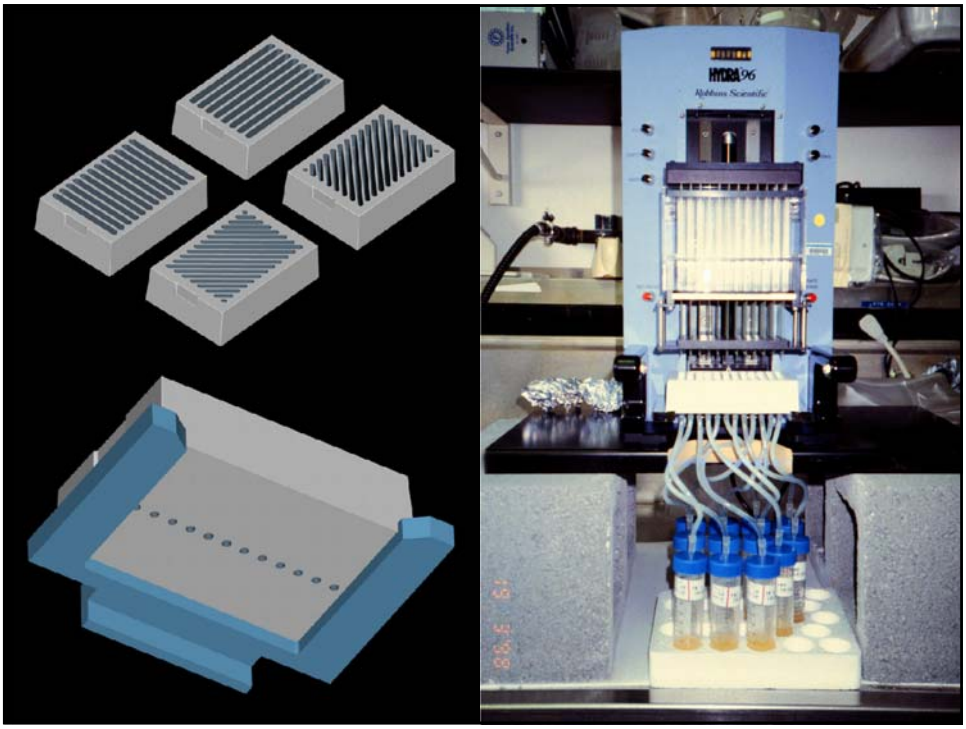
US DOE Joint Genome Institute

DOE JOINT GENOME INSTITUTE

BAC Library Pools for PCR Screening

- 8 dimensional single tier design
- pools consist of 2 sets each of plate pools, row pools, column pools and diagonal pools
- pools are constructed using Robbins Hydra and appropriate manifolds





JGI
DOE JOINT GENOME INSTITUTE

BAC end walking

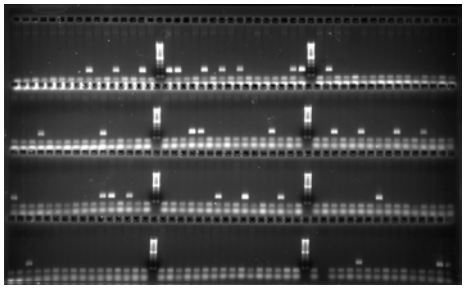
Clones adjacent to gaps

↓

BAC end sequences

↓

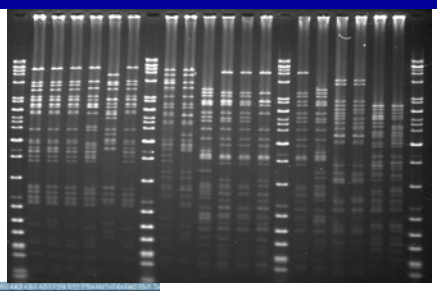
Design PCR primer and screening



Agarose gel image for screening a single STS against the single 376 pools. Marker lanes are included at 8 positions on the gel.

Restriction map assembly

All BACs are fingerprinted with EcoR I.



Restriction maps Are constructed Using GRAM

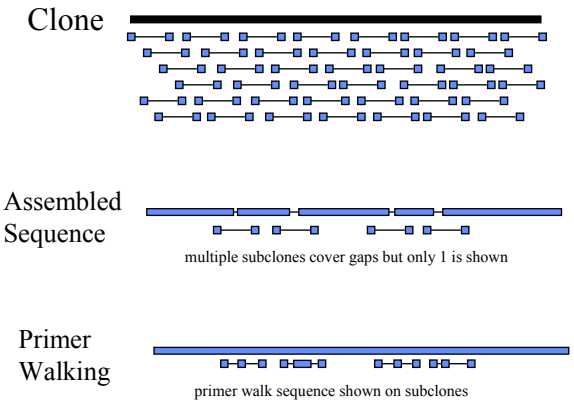
Minimal tiling set clones are selected for sequencing.

Double End Plasmid Sequencing

5X double end plasmid sequencing of 3 kb subclones

98% sequence coverage
100% subclone coverage

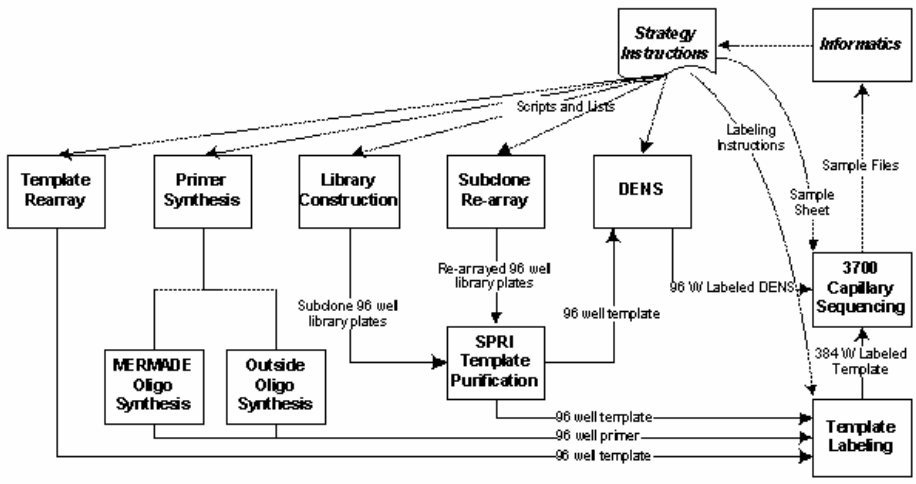
100% sequence coverage

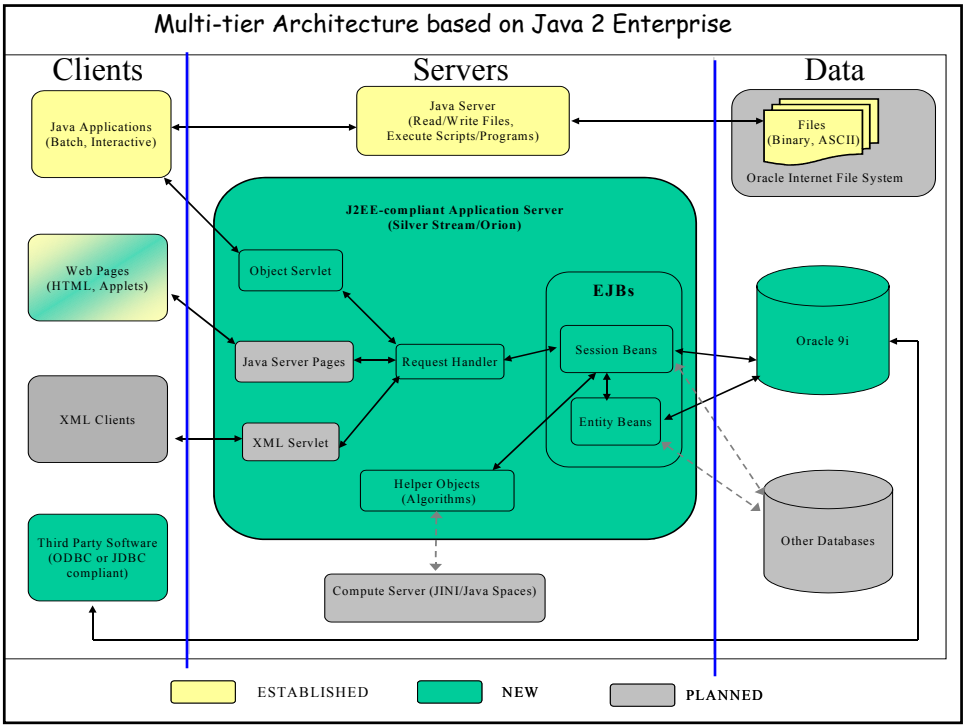


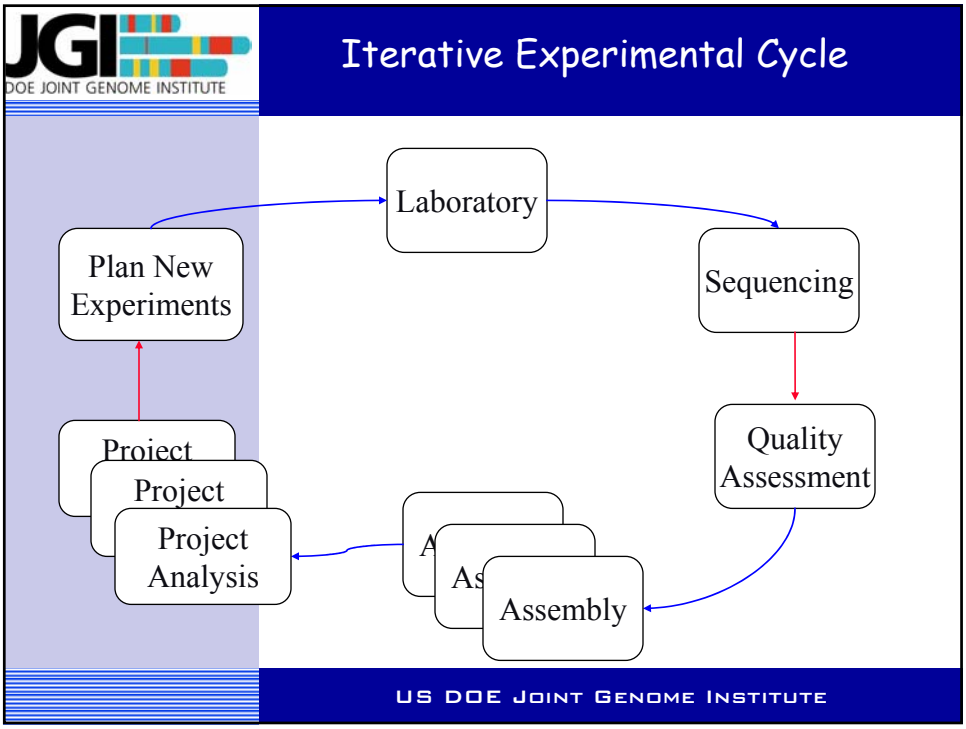
Whole Chromosome Finishing

- All BAC projects for the chromosome are considered simultaneously.
- Sequence data is merged from overlapping projects.
- A finishing target in an overlapping region is considered once.
- An Oracle database maintains records of all finishing reactions for the whole chromosome and tracks these through the finishing process.

Informatic Automation Drives all steps of Finishing



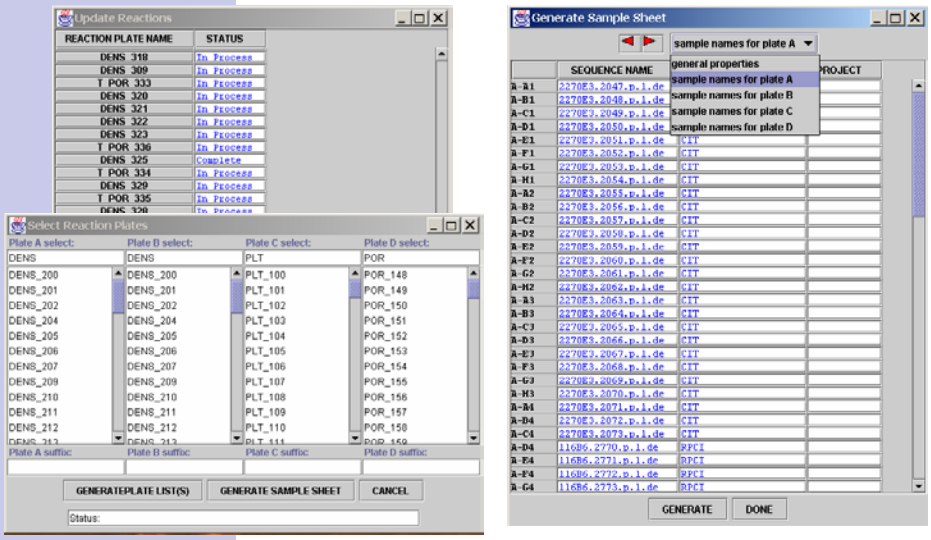

Multi-tier Architecture based on Java 2 Enterprise
Clients
Servers
Data
Java Applications (Batch, Interactive)
Java Server (Read/Write Files, Execute Scripts/Programs)
Files (Binary, ASCII)
Oracle Internet File System
J2EE-compliant Application Server (Silver Stream/Orion)
Web Pages (HTML, Applets)
Object Servlet
EJBs
Java Server Pages
Request Handler
Session Beans
Oracle 9i
XML Clients
Entity Beans
XML Servlet
Helper Objects (Algorithms)
Other Databases
Third Party Software (ODBC or JDBC compliant)
Compute Server (JINI/Java Spaces)
ESTABLISHED
NEW
PLANNED


JGI
DOE JOINT GENOME INSTITUTE
Iterative Experimental Cycle
Laboratory
Plan New Experiments
Sequencing
Quality Assessment
Project
Project
Project Analysis
Assembly
US DOE Joint Genome Institute

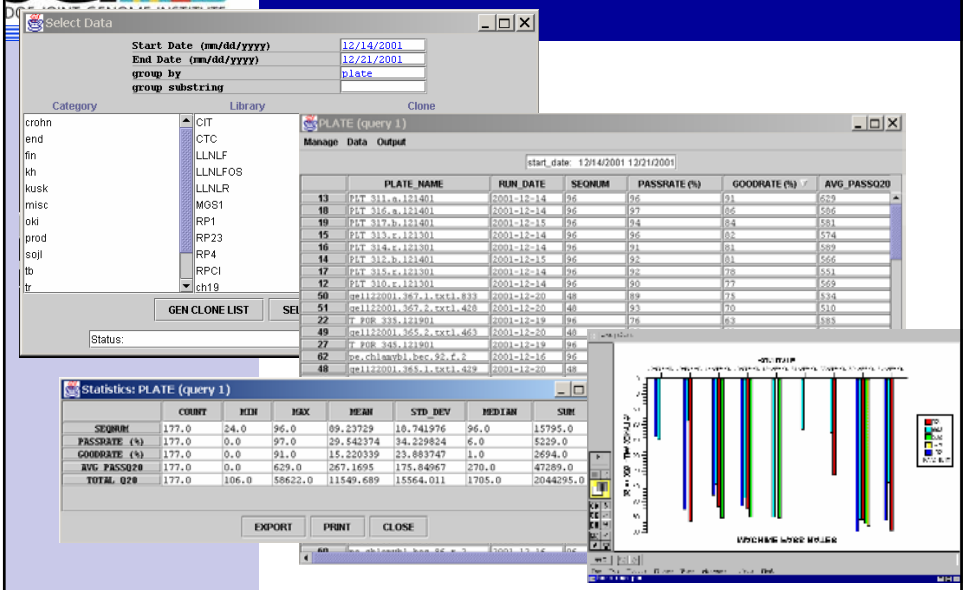
Generation of Finishing Reactions

- Primarily batch-oriented with GUI interface as necessary
- Create reactions: dp, dgtp, custom primers, dens, enhancer, shatter libraries
- Group reactions into plates by priority, source plate, primer temperature, etc
- Generate appropriate robot files

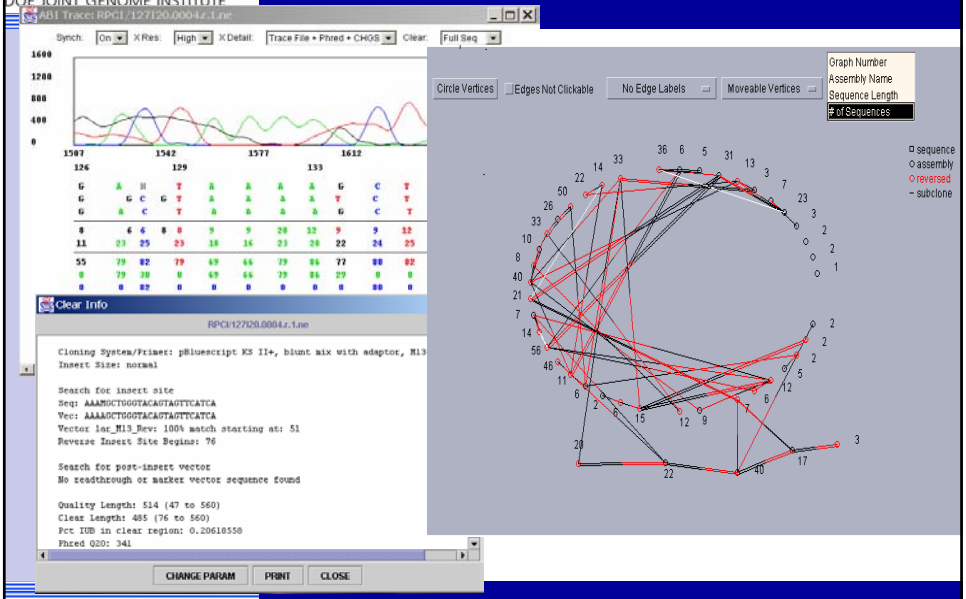
Complete Project Management



Sequence Quality and Reporting



Sequence Assembly and Analysis



Finishing Strategy

Finishing Reactions

- Big Dye reactions sequenced by ABI 3700
- dGTP Big Dye terminator reactions
- Invitrogen SequencerRx reaction enhancer or other enhancer
- Shatter libraries
- Alternate reaction chemistries
- Li-Cor infrared Global IR2 platform

- **Combine Long Reads, Redos & Primer Walks**
- **Targets & Reactions**
  - **Chemistry/Strategy Based on Desired Read Length**
  - **Geometry First for Strands**
  - **Quality Second**
  - **Priority and Alternate Reactions**
  - **Database Tracking for Efficient Rearray**

Primer Walks

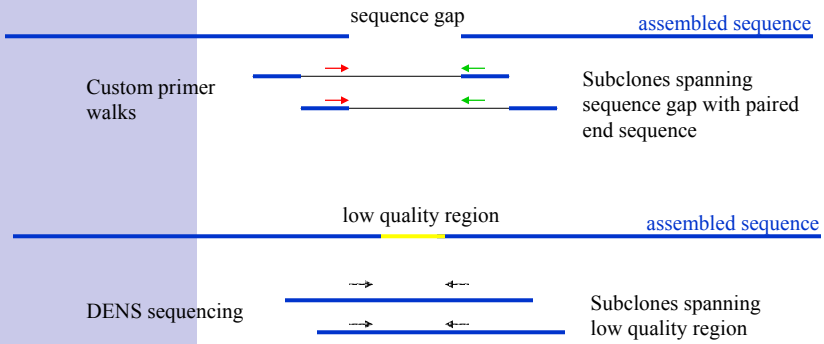
Custom Primers

- (Mermade) to Close Gaps
  - Automation for 2 Subclones/Primer
    - Q BOT Correlated Rearray
    - Half-plate Hydra Liquid Primer Handling
  - Database Generates and Tracks Primer Rxn's

DENS

- (Octomer library) for Sequence Quality Improvement
  - Database Tracks Octomer Usage and Stat's

Custom Primer Walks and DENS



DENS (Differential Extension with Nucleotide Subsets)

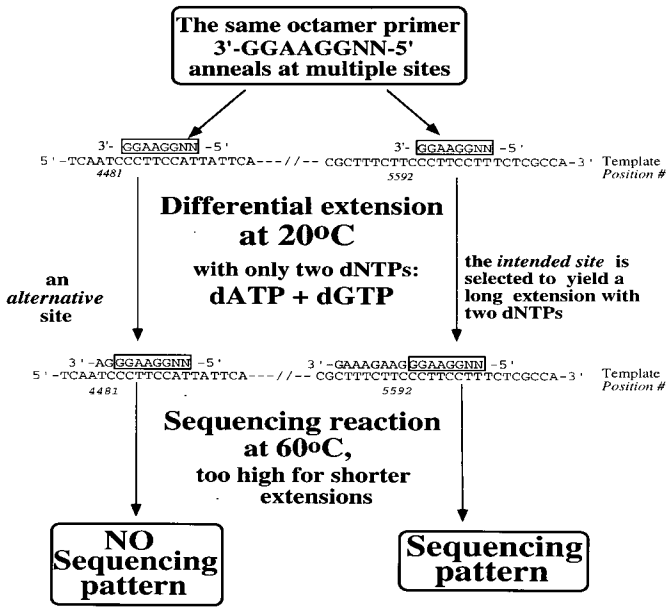
Primer walking without custom primer synthesis

NNAGGCTA

Raja et al., NUCLEIC ACIDS RESEARCH, 25:800-805 (1997) Zevin-Sonkin et al., DNA SEQUENCE, 10:245 (1999)

DENS mechanism

DENS works by converting a short primer (selected from a pre-synthesized library of 1440 double-degenerate octamers) into a long one on the template at the *intended site only*



JGI DOE JOINT GENOME INSTITUTE

Features of Finishing Pipeline

- 96 well Q-Bot clone re-array
- 96 well SPRI template prep
- 96 well MerMade oligonucleotide synthesis
- 384 well sequencing reaction using 1/6X reaction and 96 channel liquid handling
- ABI 3700 96 capillary format and automatic loading

Automation for Subclone Re-array

- Genetix Q-Bot
- Custom software to permit skips for controls, duplicating wells and flexible positioning.
- Database initiation and tracking
- 384 well compatible
- 60 re-arrayed destination plates per week

SPRI DNA Preps

- **Up to 240 96 well plates per week**
- **On universal primer, terminator chemistry, template produces pass rates of ~90% and read lengths > 650 BP**

Template Labeling

- Up to 80 384 well plates labeled per week
- Over 5,000 plates labeled per year
- Four 96 well template plates are merged to one 384 well plate
- Chemistries
  - — Custom primer or universal, Big Dye terminator
  - — Big Dye primer
  - — Universal or custom primer, dGTP Big Dye terminator

MerMade Oligonucleotide Synthesis

- Up to 20 plates synthesized per week with two MerMades
- 96 well synthesis format
- Purity confirmed with MALDI TOF Mass Spec and gel electrophoresis
- Upgraded chemical ventilation system
- Developing protocol for Universal CPG to simplify synthesis plate

Sequencing Instrumentation

- Up to 200 ABI 3700 runs per week
- Five ABI 3700 capillary sequencers
- Five ABI 377 slab gel sequencers
- One Amersham MegaBACE 1000 capillary sequencer

Human Chromosome 16 January 29, 2002

Chr16 size

Cytogenetic estimate
98 Mb total
89 Mb euchromatin

Celera Scaffolds
81 Mb

TPF Unique restfrags
77.6 Mb

Finished TPF Clones		
Total	63.5 Mb	462 Clones#
Unique	~52 Mb	(~58%)
In Finishing* (at LANL):	37.22 Mb	216 BACs
In Drafting at PGF*		
Depth (>6x):	4.77 Mb	30 BACs
In RCA (0x):	2.40 Mb	15 BACs
Glycerol (0x):	0.67 Mb	8 BACs & Cosmids

*Mb are sums of total clone sizes (not unique)

Clone Gaps (type-3):	13
Estimated remaining BACs:	18

#Cosmids, BACs, P1's, PACs and YAC

LANL Finishing Progress To Date

Month	# Clones	Total	Unique
1997	2	70,679	70,679
1998	43	2,304,776	2,031,393
1999	9	1,279,964	1,273,078
2000	20	3,376,785	3,210,578
Jan-Jul	9	1,589,347	1,269,293
Aug'01	13	2,142,612	1,613,169
Sep'01	23	4,260,548	3,431,062
Oct'01	31	4,899,475	3,372,520
Nov'01	24	3,865,978	2,894,949
Dec'01	25	3,809,468	2,269,030
Jan'02	65	10,857,259	~7,500,000
Total	264	38,456,891	~29,000,000

Who Drafted What LANL Finished?

	Draft / Finish Center / Center
17.4 Mb	PGF/LANL
15.1 Mb	LANL/LANL
5.9 Mb	G5-PGF/LANL

77.7% Unique
(through Dec'01)

Human Chromosome 16 TPF

TPF Map Status

731 total clones
32 clones w/o accessions (9 for PCR only)
13 type-3 gaps
No redundancies with other chromosomes

Clone type	Finished	Drafted	In draft	Total	% Finished
RP11 BAC	250	176	9	435	57.5%
Caltech BAC	115	57	17	189	60.8%
Cosmid	89	4	0	93	95.7%
PAC	5	3	0	8	62.5%
P1	4	0	0	4	100.0%
YAC	1	0	0	1	100.0%
PCR Fragment	0	0	1	1	0.0%
Total	464	240	27	731	63.5%

JGI DOE JOINT GENOME INSTITUTE

Human Chromosome 16 TPF

The TPF (tiling path file) is the minimal tiling set of clones that provides complete coverage of the chromosome.

Redundancy estimated as follows:
89 Mb chr16
-2.1 Mb gap
= 86.9 Mb target
108.08 Mb coverage/
87.9 Mb target = 1.244

TPF Statistics

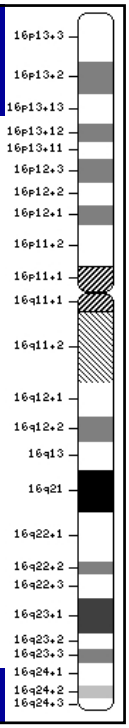
Clone distribution:	Unique Coverage:
437 RPCI-11 BACs	32.0 Mb p-arm
189 CalTech BACs	45.6 Mb q-arm
93 Cosmids	6 p-arm gaps
7 PACs	7 q-arm gaps
4 P1's	
1 YAC	

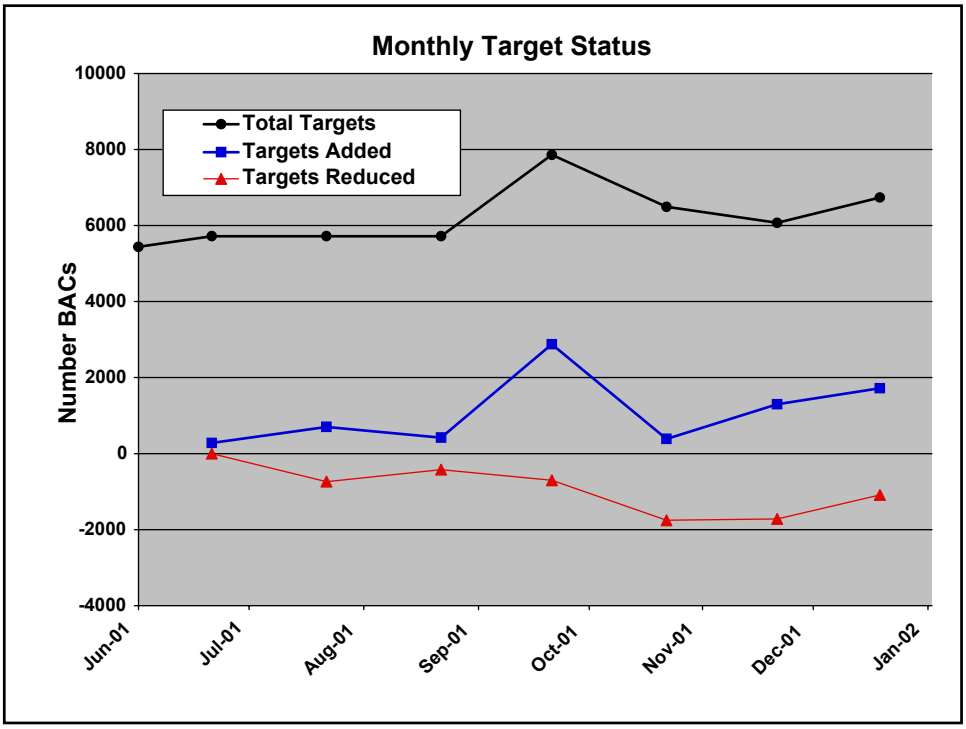
Total clone size	108,079,774 bp
Total gap size	2,100,000 bp
Estimated redundancy in TPF	1.244

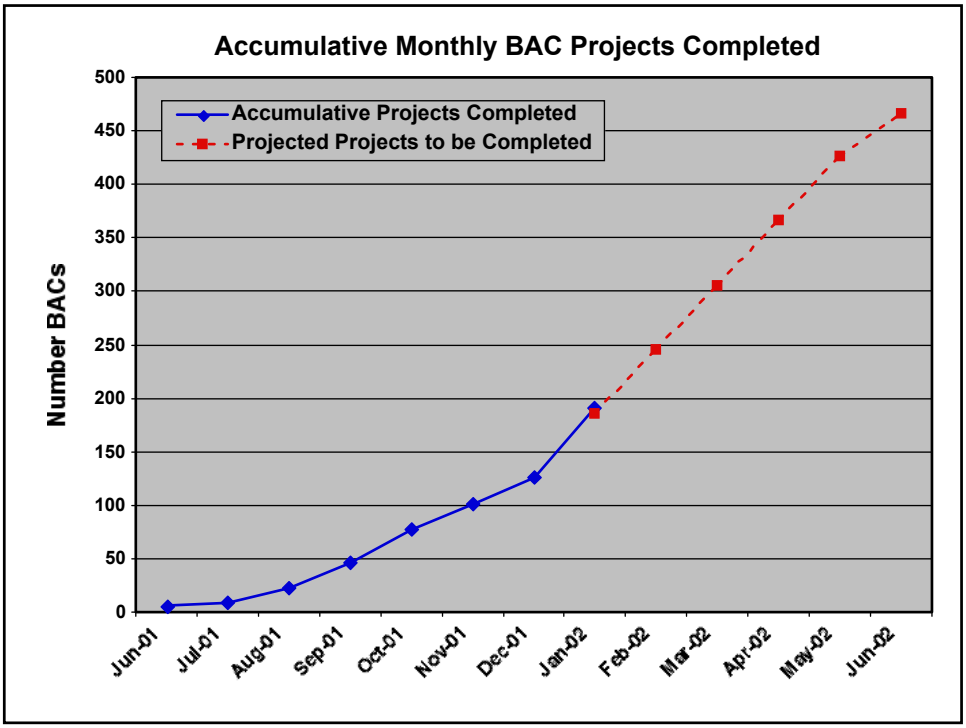
Major Finishers on TPF (total of clone sizes)

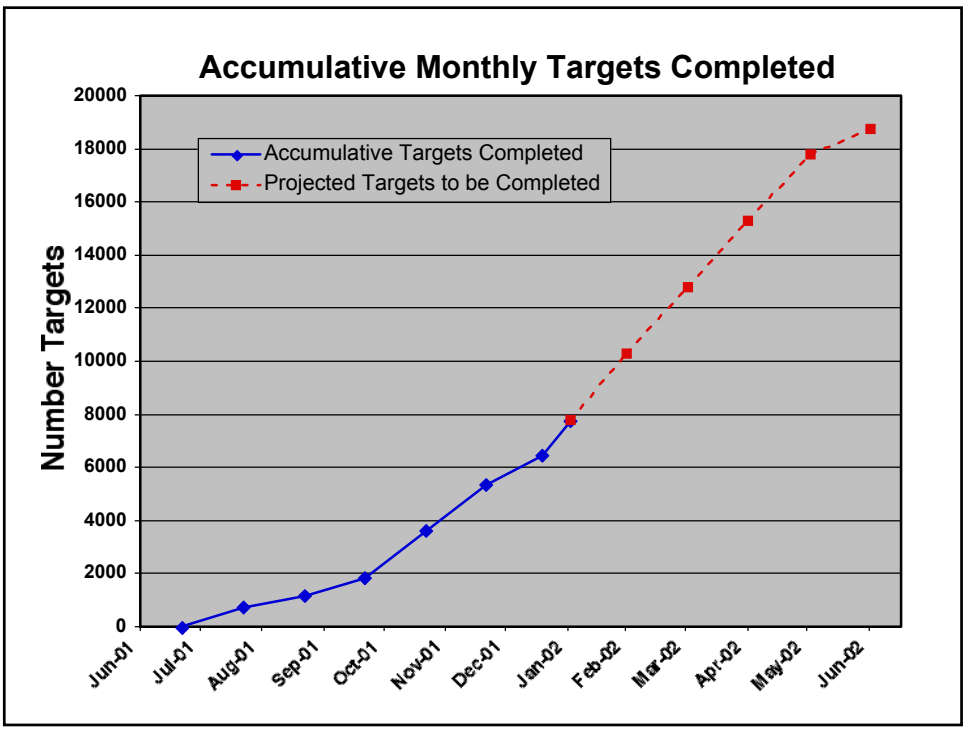
LANL	38,456,891
Stanford	13,157,867
TIGR	9,674,910
Sanger	1,442,639
Wash U	1,156,547

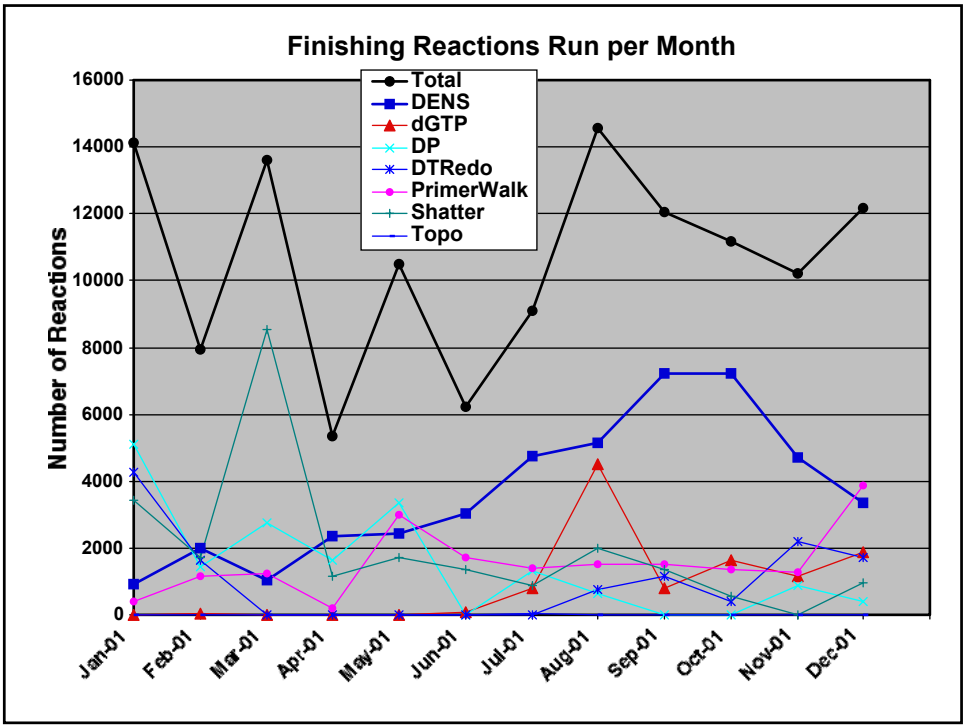
US DOE Joint Genome Institute






Accumulative Monthly BAC Projects Completed
Number BACs
Accumulative Projects Completed
Projected Projects to be Completed
500
450
400
350
300
250
200
150
100
50
0
Jun-01
Jul-01
Aug-01
Sep-01
Oct-01
Nov-01
Dec-01
Jan-02
Feb-02
Mar-02
Apr-02
May-02
Jun-02


Accumulative Monthly Targets Completed
Number Targets
Accumulative Targets Completed
Projected Targets to be Completed
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
Jun-01
Jul-01
Aug-01
Sep-01
Oct-01
Nov-01
Dec-01
Jan-02
Feb-02
Mar-02
Apr-02
May-02
Jun-02

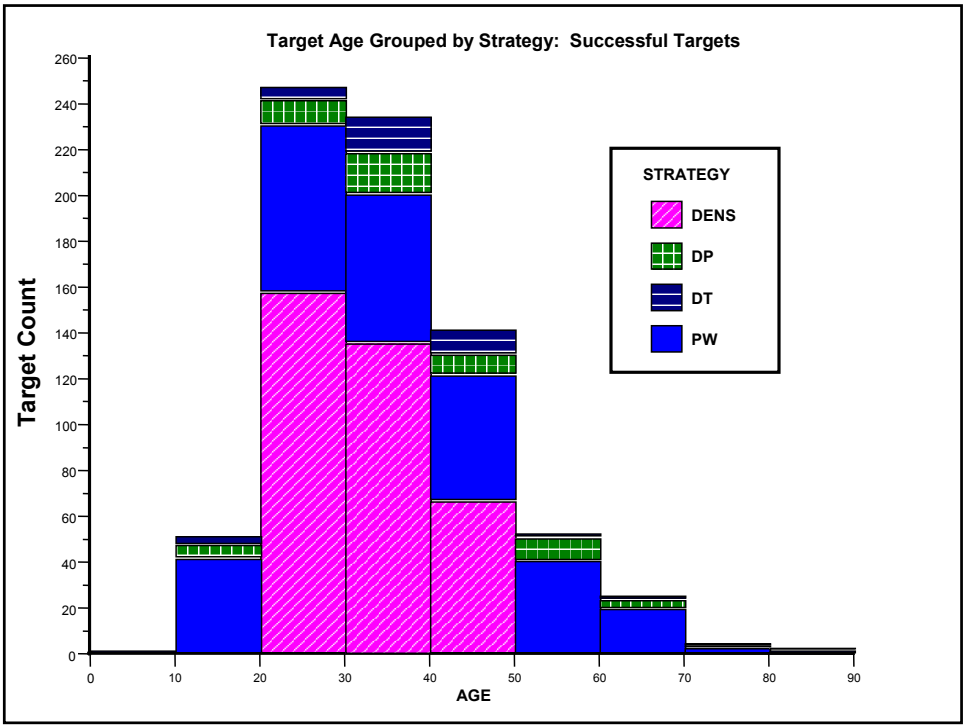


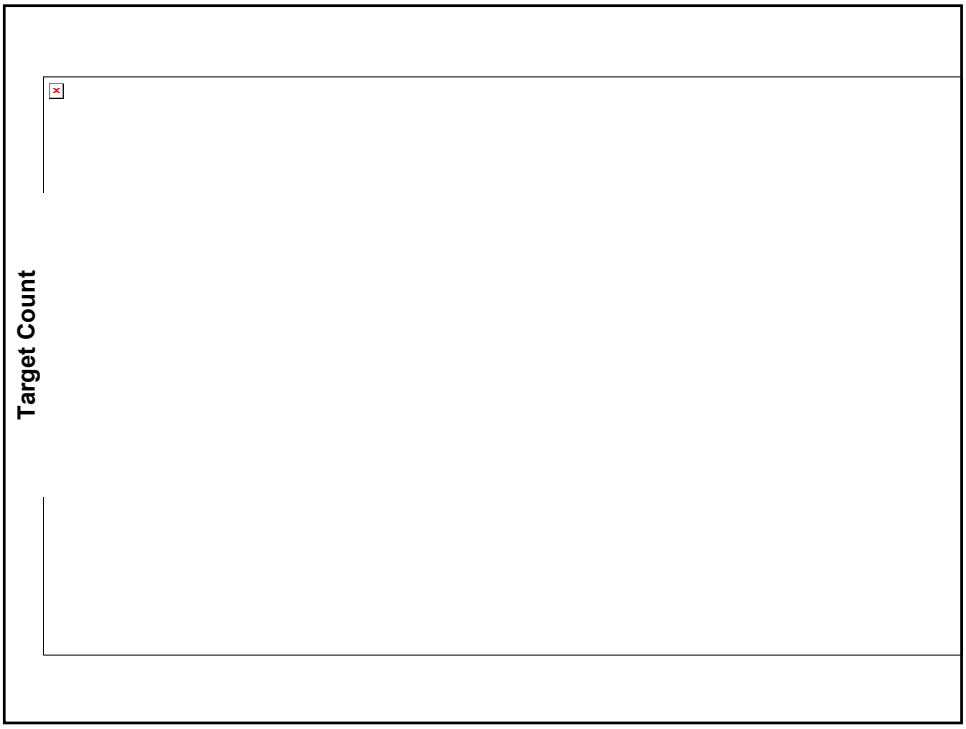
DOE JOINT GENOME INSTITUTE

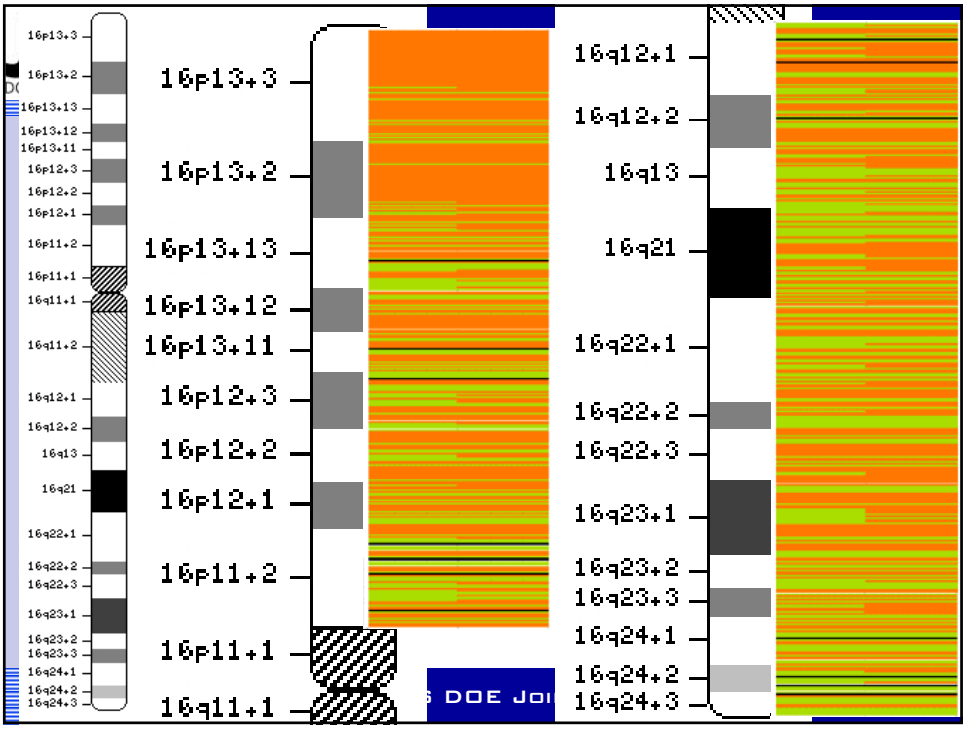
Finishing Reads per Kb

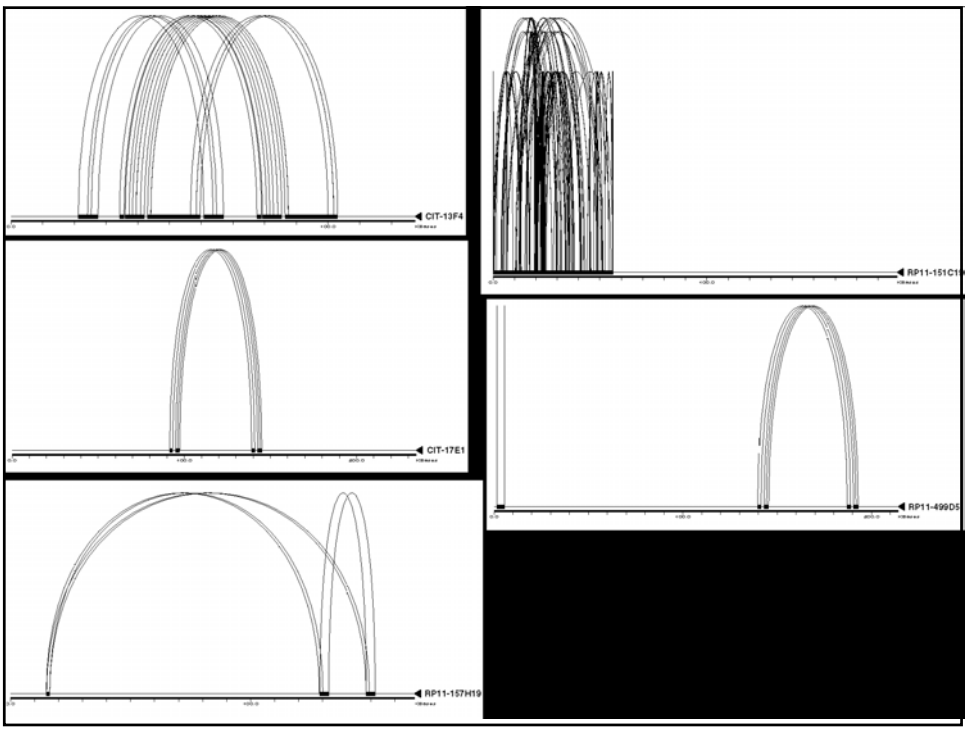
Draft/Finish	Min.	Avg.	Max.
LANL/LANL	0.3	3.4	7.2
JGI/LANL	0.1	3.2	12.0
G5/JGI/LANL	0.1	0.8	1.5

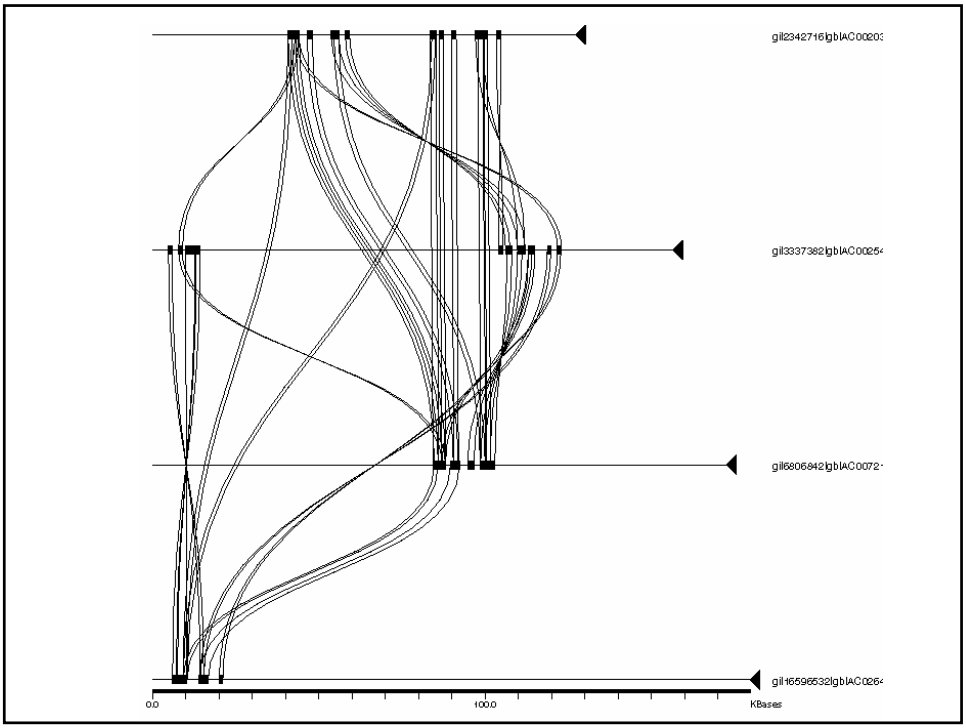
(Data from ~ 100 BAC projects finished in 2000-2001)

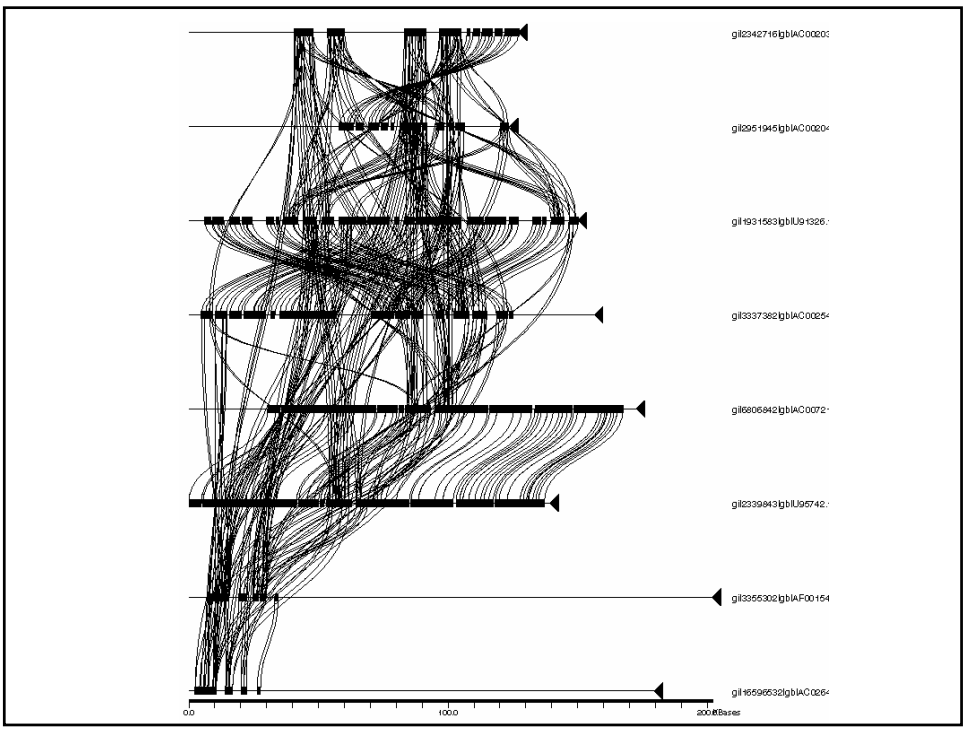

Target Age Grouped by Strategy: Successful Targets
Target Count
260
240
220
200
180
160
140
120
100
80
60
40
20
0
0
10
20
30
40
50
60
70
80
90
AGE
STRATEGY
DENS
DP
DT
PW


Target Count


16p13.3
16p13.2
16p13.13
16p13.12
16p13.11
16p12.3
16p12.2
16p12.1
16p11.2
16p11.1
16q11.1
16q11.2
16q12.1
16q12.2
16q13
16q21
16q22.1
16q22.2
16q22.3
16q23.1
16q23.2
16q23.3
16q24.1
16q24.2
16q24.3
16p13.3
16p13.2
16p13.13
16p13.12
16p13.11
16p12.3
16p12.2
16p12.1
16p11.2
16p11.1
16q11.1
16q12.1
16q12.2
16q13
16q21
16q22.1
16q22.2
16q22.3
16q23.1
16q23.2
16q23.3
16q24.1
16q24.2
16q24.3


CIT-13F4
CIT-17E1
RP11-157H19
RP11-151C1
RP11-499D5


gil2342716|gb|AC00203
gil3337382|gb|AC00254
gil6806842|gb|AC0072
gil16596532|gb|AC0264
0.0
100.0
KBases


gil2342716|gb|AC00203
gil2951945|gb|AC00204
gil1931583|gb|U91326.
gil3337382|gb|AC00254
gil6806842|gb|AC0072
gil2339843|gb|U95742.
gil3355302|gb|AF00154
gil16596532|gb|AC0264
0.0
100.0
200.0KBases

Acknowledgements

DOE Joint Genome Institute and Los Alamos Center
for Human Genome Studies

- Trevor Hawkins
- Sue Lucas
- Joe Monforte

Marv Frazier

- Mark Mundt
- David Bruce
- Judith Cohn
- Levy Ulanovsky
- Larry Deaven

Acknowledgements

US DOE Joint Genome Institute